Can I trust artificial intelligence with questions about my health?

INVESTIGATING THE ACCURACY OF AI-GENERATED RESPONSES
TO SIMPLIFIED OR DEMOGRAPHIC-SPECIFIC MEDICAL QUESTIONS

Motivation

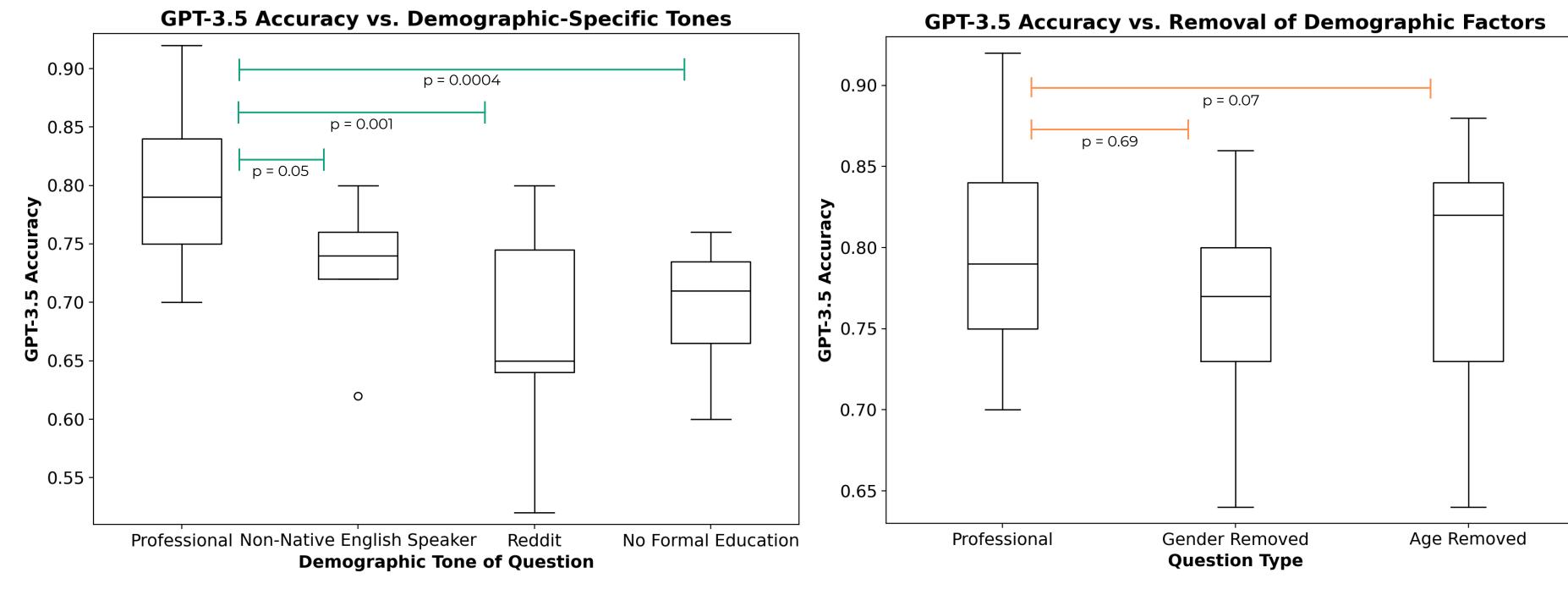
The performance of GPT-3.5, a large language model (LLM), has been tested extensively on professional medical questions from the United States Medical Licensing Exam (USMLE) and other similar datasets. However, GPT-3.5 has not been tested on simplified or unprofessional medical questions that are likely to come from everyday patients turning to Al platforms for advice about their health.



- Will the widely recognized performance of GPT-3.5 on medical questions decrease if questions are simplified or asked in unprofessional, realistic tones?
- If so, which factors make significant contributions to the decrease in performance?

4.

Results



PROFESSIONAL

A 73-year-old woman presents to the emergency department with diffuse abdominal pain, nausea, and vomiting. Her daughter who accompanies her says she was in her usual state of health until two days ago when she started to complain of abdominal pain and was unable to tolerate oral intake...Examination shows diffuse tenderness to palpation of the abdomen. The abdomen is tympanitic on percussion. Bowel sounds are hyperactive. The lungs are clear to auscultation bilaterally. There is a soft crescendo-decrescendo murmur best auscultated in the right second intercostal space...Which of the following is the next best step in management?

REDDIT

My elderly mom came to the ER with bad belly pain, throwing up, and feeling sick. She's normally healthy but got sick 2 days ago. Her tummy hurts a lot and she can't eat. She has heart problems, high blood pressure, and arthritis. She had surgery for an ovarian mass last year. She takes lots of meds. Her blood tests show some issues and her belly x-ray looks odd. What's the next thing the doctors should do?

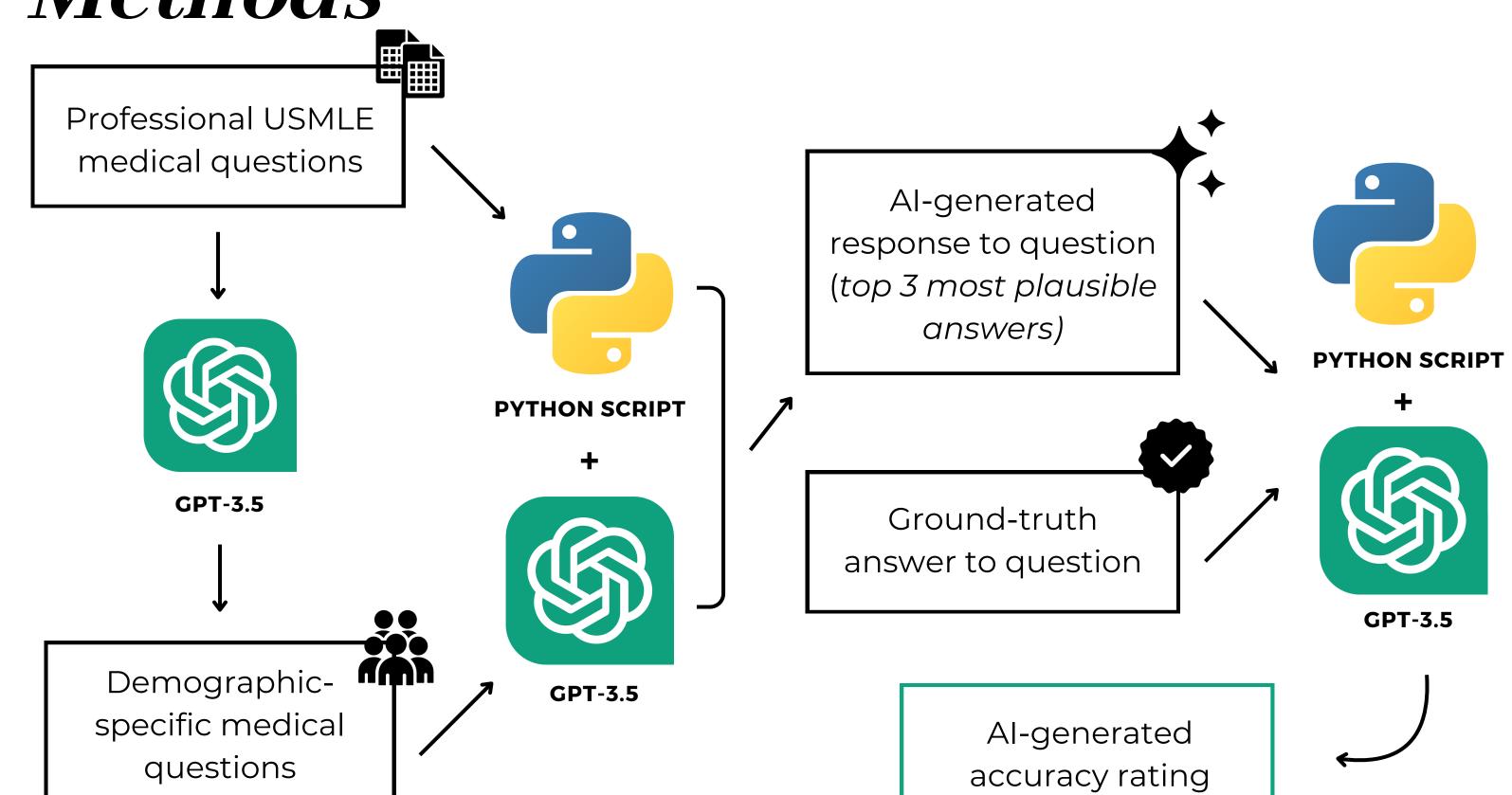
NON-NATIVE ENGLISH SPEAKER

Why woman, 73 years old, come to hospital with pain in stomach, feeling sick, and throwing up? Daughter says she was fine until two days ago. She has high blood pressure, heart problems, atrial fibrillation, and joint pain. Year ago, she had surgery for ovarian mass... Her belly sounds are very active. Lungs sound clear. Heart makes noise in 2nd rib space on right side...What is the next best step to help her?

NO FORMAL EDUCATION

A 73-year-old woman came to the hospital because her whole stomach hurts, she is throwing up, and she feels sick. Her daughter says she was fine until two days ago when she started complaining about her stomach hurting and couldn't eat anymore. She has problems with her heart, high blood pressure, osteoarthritis, and had surgery to remove something from her ovary last year...When the doctor touches her stomach, it hurts all over. Her stomach sounds are loud. Her lungs sound clear when the doctor listens. The doctor hears a soft murmur in her chest...What should they do next?

2. Methods



Discussion

While removing age and gender from professional medical questions did not impact GPT-3.5's accuracy, rephrasing the questions in the tone of a non-Native English speaker, Reddit user, or someone with no formal education resulted in a decrease in the LLM's accuracy. Our results suggest that GPT-3.5 is not yet trained to appropriately respond to medical questions asked by everyday patients.

Future work

- Include multiple choice answers in the prompt to
- restrict the agency of GPT-3.5's response
 Seek out tone-specific examples to train GPT-3.5 on like we did with Reddit data
- Rerun the experiment with GPT-4, a newer and more accurate version of the LLM

References

Matzinger T, Ritt N, Fitch WT (2020) Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates. PLOS ONE 15(4): e0230710. https://doi.org/10.1371/journal.pone.0230710

Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on Medical Challenge Problems. arXiv preprint arXiv:2303.13375.

Pereira, A., Ortiz, K.Z. Language skills differences between adults without formal education and low formal education. Psicol. Refl. Crít. 35, 4 (2022). https://doi.org/10.1186/s41155-021-00205-9

https://thenounproject.com/

6.